Large N Limit

As we get more data, it would be nice if the APC we compute tends toward the right answer, equation (2) of Gelman 2007. I don't care about asymptotics as much as some people, but if we don't get the right answer in the limit, that's at least a clue that we might not be doing as well as we can for smaller N. This note shows that unless we adjust the weights weighting function as we get more data, we won't have that nice property.

```
makeExampleDF <- function(N) {
  exampleDF <- data.frame(
    v=c(3,3,7,7),
    u=c(10,20,12,22)
    )[rep(c(1,2,3,4),c(.4*N,.4*N,.1*N,.1*N)),]
  exampleDF <- transform(exampleDF, v = v + rnorm(nrow(exampleDF), sd=.001))
  return(exampleDF)
}
```

Just as in the note "Normalizing Weights", the APC should be:

$$.8\delta_u(10 \to 20, 3, f) + 0.2\delta_u(12 \to 22, 7, f) = (.8)(3) + (.2)(6) = 3.8$$

We get almost the same APC with 300 data points as 100:

```
get_apc(function(df) return(df$u * df$v), makeExampleDF(100), u="u", v="v")
```

```
## [1] 3.854
```

```
get_apc(function(df) return(df$u * df$v), makeExampleDF(300), u="u", v="v")
```

```
## [1] 3.845
```

If we're looking at one value for v, $v = v_0$, the tradeoff in determining the weights is:

1. v's closer to v_0 will do a better job representing the distribution of u conditional on $v = v_0$
2. but if too few v's get too much of the weight, our estimate for the conditional distribution of u will be too noisy

As we get more data, we can afford to put more weight on closer v, because (2) becomes less of a problem. A couple ideas are:

- With the weights as $\frac{1}{k+d}$ (where now $k = 1$ and d is the Mahalanobis distance), we could scale k down as N goes up.
- Or use the weights we are now, except we drop (set the weight to 0) for all but the closest $s(N)$ points to each v. The function s needs to increases with N, but not as fast as N, e.g. maybe $s(N) = sqrt(N)$ probably works. – this means we're always decreasing bias (sampling from closer to the right v) and also decreasing variance (more samples) as N increases.